

**Genomic Signatures of Historical Allopatry and Ecological Divergence in an Island Lizard**

Análise de sequências biológicas

Licenciatura de Bioinformática

30/06/2023

João Ribeiro Nº202101169

Pedro Pacheco Nº202100957

Índice

[1.Introdução 4](#_Toc139042227)

[1.1.Problema biológico e Objetivos 4](#_Toc139042228)

[1.2.Métodos Utilizados 4](#_Toc139042229)

[1.3.Conclusão 5](#_Toc139042230)

[2.Objetivos 6](#_Toc139042231)

[3.Materiais e métodos utilizados 6](#_Toc139042232)

[4.Resultados 6](#_Toc139042233)

[5.Discussão 6](#_Toc139042234)

[6.Bibliografia 8](#_Toc139042235)

# 1.Introdução

## 1.1.Problema biológico e Objetivos

A técnica conhecida como Sequenciamento de DNA Associado a Sítios de Restrição (RAD-seq) apresenta-se como uma alternativa viável e economicamente vantajosa em comparação ao sequenciamento de genoma completo para estudos populacionais envolvendo uma amostragem extensa de indivíduos. Essa abordagem possui uma vantagem adicional ao ser aplicável a espécies não modelo (Brown et al., 2016). Enquanto que análises tradicionais de padrões biogeográficos, associados à alopatria histórica, costumam se basear no sequenciamento do DNA mitocondrial e em um número limitado de marcadores nucleares, o RAD-seq amplia significativamente a proporção do genoma analisado, permitindo uma avaliação mais minuciosa e robusta dos processos subjacentes à divergência populacional. Além disso, o RAD-seq auxilia na identificação de regiões genómicas de interesse, o que pode fornecer insights valiosos (Brown et al., 2016).

O lagarto Gallotia galloti é encontrado em toda a Ilha Canária de Tenerife, exibindo variação morfológica significativa ao longo da ilha relativamente grande para uma ilha tão pequena como esta (2 034 km2) (Brown et al., 2016). A vegetação nas encostas voltadas para o norte é mais úmida, com densa floresta nublada e outras plantas, enquanto as encostas voltadas para o sul são mais áridas, com comunidades adaptadas ao clima seco (THORPE & BROWN, 1989). A cor dorsal dos machos adultos e o tamanho do corpo estão relacionados a esses habitats distintos (Thorpe & Brown, 1989; THORPE & BROWN, 1989). Estudos mostraram que há diferenças genéticas entre os lagartos das regiões norte e sul, indicando uma possível especiação incipiente. No entanto, outros aspetos da morfologia, como a escala, apresentam padrões geográficos diferentes (Brown et al., 2006; Thorpe & Baez, 1987), sugerindo origens complexas para a diversidade morfológica.

A identificação completa das causas subjacentes à variação intra-ilha em G. galloti é um desafio complexo devido à presença de duas linhagens distintas de DNA mitocondrial (mtDNA), que divergiram aproximadamente a 0,8 milhões de anos atrás(Brown et al., 2006). Independentemente do mecanismo real envolvido, a divergência alopátrica fornece uma explicação alternativa à hipótese de diferenças nas pressões de seleção *in situ*. No entanto, essa estruturação geográfica diverge do padrão de variação morfológica descrito anteriormente e tem levado à sugestão de que os padrões geográficos do mtDNA e a variação morfológica possuem causas independentes (Thorpe et al., 1996; Thorpe & Richard, 2001)**.**

O objetivo desta pesquisa consistiu em avaliar a aplicabilidade da abordagem RAD-seq na distinção entre as hipóteses de seleção divergente entre habitats e isolamento histórico como explicações para a variação morfológica correlacionada ao habitat em G. galloti (Brown et al., 2016).

## 1.2.Métodos Utilizados

Neste ponto iremos apenas referenciar os processos de obtenção dos dados e o seu processamento até a obtenção do PCA que é o foco principal deste relatório.

Foram coletadas amostras de G. galloti em nove locais em Tenerife, que serão referidos como "populações amostradas". A amostragem foi planejada para incluir pares próximos de populações nas encostas norte e sul, cobrindo áreas ocupadas por ambas as linhagens de mtDNA. As amostras de tecido foram obtidas de todos os indivíduos e armazenadas em DNAgard®. O DNA genómico completo foi extraído usando um kit de sangue e tecido Qiagen DNeasy.

O RAD-seq foi realizado em 148 amostras, representando 135 indivíduos, para a preparação das bibliotecas RAD SbfI, o DNA digerido foi ligado a adaptadores P1 contendo códigos de barras individuais. Os fragmentos foram sonicamente cortados para uma faixa de tamanho de 300 a 600 pb. Adaptadores P2 foram ligados às extremidades cortadas (Brown et al., 2016). As bibliotecas foram amplificadas por PCR e sequenciadas utilizando uma célula de fluxo Illumina HiSeq 2000. Seguidamente foram analisadas as sequencias usando a abordagem RADmapper, baseada no *pipeline* Stacks RAD (Catchen et al., 2013).

Realizaram-se análises genómicas após a exclusão de amostras replicadas e seis indivíduos não replicados, de cinco populações distintas, devido a uma baixa taxa de mapeamento das leituras para os tags montados. A partir dessa seleção, um conjunto inicial de SNPs foi obtido, sendo então submetido a um processo de filtragem para identificar um grupo mais restrito de SNPs altamente confiáveis (Brown et al., 2016).

E por fim foram realizadas análises multivariadas dos dados de SNPs no ambiente R (Jombart & Bateman, 2008). A Análise de Componentes Principais (PCA) foi empregada para examinar a diferenciação global entre os indivíduos, independentemente de sua população de origem (Brown et al., 2016).

## 1.3.Conclusão

As análises genómicas indicam que tanto a divergência ecológica como a alopatria antiga se combinaram para moldar a variação genética atual. No entanto, as principais diferenças fenotípicas parecem ser determinadas por diferenças de habitat, o que sugere que os padrões diferenciais de seleção estão a conduzir a divergência e a especiação potencial, em vez do isolamento histórico da população (Brown et al., 2016).

Isso nos leva a sugerir que padrões seletivos distintos têm desempenhado um papel crucial na condução da divergência genética e, potencialmente, na ocorrência de especiação, em vez de apenas fatores históricos isolacionistas.

A especiação ocorre quando populações de uma mesma espécie se separam e evoluem de forma independente, levando ao surgimento de novas espécies. Embora o isolamento geográfico tenha sido tradicionalmente considerado o principal fator para a especiação, evidências recentes mostram que outros mecanismos, como a seleção natural diferencial, também podem desempenhar um papel importante.

Ao considerar as diferenças fenotípicas observadas, relacionadas às características visíveis de um organismo, como forma, cor e comportamento, fica evidente que elas estão correlacionadas com as adaptações aos diferentes habitats. Por exemplo, em ambientes com recursos alimentares distintos, diferentes características físicas e fisiológicas podem ser favorecidas pela seleção natural. Assim sendo essas diferenças adaptativas podem levar à formação de grupos distintos dentro da mesma espécie, que podem eventualmente se tornar espécies separadas.

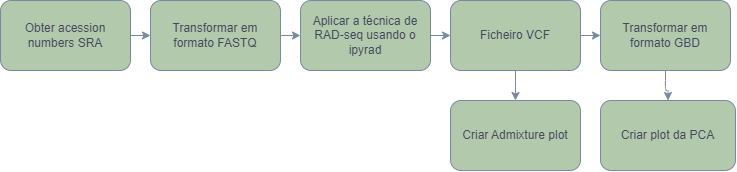
# 2.Objetivos

# 3.Materiais e métodos utilizados

Inicialmente, obtivemos um arquivo no formato CSV contendo os números de acesso de cada arquivo no formato SRA por meio do site do NCBI. Em seguida, utilizamos um programa disponível no [GitHub](https://github.com/joao-ribeirooo/Assigment-2/blob/main/bioproject_download_and_compress.sh) para ler esses números de acesso dos arquivos SRA, convertendo-os para o formato FASTQ e, posteriormente, realizamos a compressão individual de cada arquivo por meio de um processo de arquivamento em formato zip. Consequentemente, utilizamos outro programa também disponível no [GitHub](https://github.com/joao-ribeirooo/Assigment-2/blob/main/bioproject_file_processing.sh) para abrir cada arquivo FASTQ, selecionando apenas as 500.000 primeiras linhas, e, em seguida, realizamos a compressão desses arquivos, armazenando-os em um diretório específico.

Posteriormente, utilizamos o software ipyrad (versão 0.9.90) para realizar a técnica de RAD-seq. Para isso, criamos um arquivo de parâmetros contendo 29 parâmetros distintos necessários para a execução adequada do ipyrad. Durante a configuração desse arquivo, efetuamos modificações específicas nos parâmetros 4, 8 e 27. No parâmetro 4, indicamos o diretório onde estavam localizados os arquivos FASTQ após o processo de seleção das 2.000.000 primeiras linhas. No parâmetro 8, alteramos a enzima de restrição para SbfI, a qual foi utilizada no artigo em questão. Por fim, no parâmetro 27, realizamos ajustes para obter todos os tipos de formatos de saída disponíveis.

Com todas as configurações definidas, executamos o software ipyrad utilizando o seguinte comando: **ipyrad -p params-simdata.txt -s 1234567 -c 6**. Nesse comando, o parâmetro "-p" representa o caminho para o arquivo de parâmetros do ipyrad, o parâmetro "-s" especifica os passos a serem executados, variando de 1 a 7, e o parâmetro "-c" indica o número de núcleos de processamento utilizados durante a execução.



# 4.Resultados

# 5.Discussão

# 6.Bibliografia

Brown, R. P., Hoskisson, P. A., Welton, J. H., & Báez, M. (2006). Geological history and within-island diversity: a debris avalanche and the Tenerife lizard Gallotia galloti. *Molecular Ecology*, *15*(12), 3631–3640. https://doi.org/10.1111/J.1365-294X.2006.03048.X

Brown, R. P., Paterson, S., & Risse, J. (2016). Genomic Signatures of Historical Allopatry and Ecological Divergence in an Island Lizard. *Genome Biology and Evolution*, *8*(11), 3618–3626. https://doi.org/10.1093/GBE/EVW268

Catchen, J., Hohenlohe, P. A., Bassham, S., Amores, A., & Cresko, W. A. (2013). Stacks: an analysis tool set for population genomics. *Molecular Ecology*, *22*(11), 3124–3140. https://doi.org/10.1111/MEC.12354

Jombart, T., & Bateman, A. (2008). adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics*, *24*(11), 1403–1405. https://doi.org/10.1093/BIOINFORMATICS/BTN129

Thorpe, R. S., & Baez, M. (1987). GEOGRAPHIC VARIATION WITHIN AN ISLAND: UNIVARIATE AND MULTIVARIATE CONTOURING OF SCALATION, SIZE, AND SHAPE OF THE LIZARD GALLOTIA GALLOTI. *Evolution*, *41*(2), 256–268. https://doi.org/10.1111/J.1558-5646.1987.TB05795.X

Thorpe, R. S., Black, H., & Malhotra, A. (1996). Matrix Correspondence Tests on the DNA Phylogeny of the Tenerife Lacertid Elucidate both Historical Causes and Morphological Adaptation. *Systematic Biology*, *45*(3), 335–343. https://doi.org/10.1093/SYSBIO/45.3.335

THORPE, R. S., & BROWN, R. P. (1989). Microgeographic variation in the colour pattern of the lizard Gallotia galloti within the island of Tenerife: distribution, pattern and hypothesis testing. *Biological Journal of the Linnean Society*, *38*(4), 303–322. https://doi.org/10.1111/J.1095-8312.1989.TB01580.X

Thorpe, R. S., & Brown, R. P. (1989). Testing hypothesized causes of within-island geographic variation in the colour of lizards. *Experientia*, *45*(4), 397–400. https://doi.org/10.1007/BF01957493/METRICS

Thorpe, R. S., & Richard, M. (2001). Evidence that ultraviolet markings are associated with patterns  of molecular gene flow. *Proceedings of the National Academy of Sciences*, *98*(7), 3929–3934. https://doi.org/10.1073/PNAS.071576798